

Course Title: Artificial Intelligence and Expert Systems
Date: 18.4.2015 (Second term)Course Code: CCE3219 3rd year
Allowed time: 1 hrs**Question No. 1****(8 marks- 1 mark for each)**

For each of the following, please circle the letter introducing the best answer.
(Check all that apply.) Explain your answer.

1. Suppose you are working on stock market prediction, and you would like to predict whether or not a particular stock's price will be higher tomorrow than it is today. You want to use a learning algorithm. Which one of the following algorithms is appropriate?

- a) Regression
- b) Classification
- c) Clustering
- d) Reinforcement learning

The correct choice b

Classification is appropriate when we are trying to predict one of a small number of discrete-valued outputs. Here, there are two possible outcomes: That the stock price goes up (which we might designate as class 0, say) or that it does not (class 1).

2. A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. In this setting, what is T?

- a) The process of the algorithm examining a large amount of historical weather data.
- b) The weather prediction task.
- c) The probability of it correctly predicting a future date's weather.
- d) None of these.

The correct choice b

The task described is weather prediction, so this is Task T.

3. Let f be some function so that $f(\theta_0, \theta_1)$ outputs a number. For this problem, f is some arbitrary/unknown smooth function (not necessarily the cost function of linear regression, so may have local optima). Suppose we use gradient descent to try to minimize $f(\theta_0, \theta_1)$ as a function of θ_0 and θ_1 . Which of the following statements are true? (Check all that apply.)

- a) If the first few iterations of gradient descent cause $f(\theta_0, \theta_1)$ to increase rather than decrease, then the most likely cause is that we have set the learning rate α to too large a value.
- b) If θ_0 and θ_1 are initialized at a local minimum, the one iteration will not change their values.
- c) No matter how θ_0 and θ_1 are initialized, so long as α is sufficiently small, we can safely expect gradient descent to converge to the same solution.

- d) Setting the learning rate α to be very small is not harmful, and can only speed up the convergence of gradient descent.

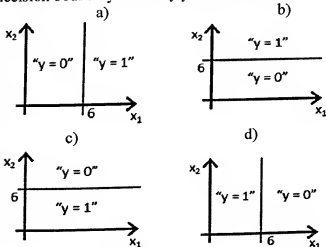
The correct choice a, b

- If alpha were small enough, then gradient descent should always successfully take a tiny small downhill and decrease $f(\theta_0, \theta_1)$ at least a little bit. If gradient descent instead increases the objective value, that means alpha is too large (or you have a bug in your code!).
 - At a local minimum, the derivative (gradient) is zero, so gradient descent will not change the parameters.
 - Depending on the initial condition, gradient descent may end up at different local optima.
 - If the learning rate is small, gradient descent ends up taking an extremely small step on each iteration, so this would actually slow down (rather than speed up) the convergence of the algorithm.
4. Suppose you have a dataset with $m=100000$ examples and $n=15$ features for each example. You want to use multivariate linear regression to fit the parameters to our data. Should you prefer gradient descent or the normal equation?
- The normal equation, since gradient descent might be unable to find the optimal θ .
 - The normal equation, since it provides an efficient way to directly find the solution.
 - Gradient descent, since it will always converge to the optimal θ .
 - Gradient descent, since $(X^T X)^{-1}$ will be very slow to compute in the normal equation.

The correct choice b

With $n=15$ features, you will have to invert a 15×15 matrix to compute the normal equation. This is a simple inversion, so the normal equation is efficient.

5. Suppose you train a logistic classifier $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$. Suppose $\theta_0 = -0.6$, $\theta_1 = 0$, and $\theta_2 = 1$. Which of the following figures represents the decision boundary found by your classifier?



The correct choice b

In this figure, we transition from negative to positive when x_2 goes from below 6 to above 6 which is true for the given values of θ .

6. You are training a classification model with logistic regression. Which of the following statements are true? Check all that apply.
- a) Introducing regularization to the model always results in equal or better performance on the training set.
 - b) Adding a new feature to the model always results in equal or better performance on the training set.
 - c) Adding a new feature to the model always results in equal or better performance on examples not in the training set.
 - d) Introducing regularization to the model always results in equal or better performance on examples not in the training set.

The correct choice b

By adding a new feature, our model must be more (or just as) expressive, thus allowing it learn more complex hypotheses to fit the training set.

7. Consider an A^* search algorithm for which $h(n) = 0$. To which of the following search algorithms is this A^* equivalent?
- a) Greedy best-first search
 - b) Depth-First Search
 - c) Uniform Cost Search
 - d) None of the above.

The correct choice c

Uniform cost search is A^* search with no heuristics

8. K-means is an iterative algorithm, and two of the following steps are repeatedly carried out in its inner-loop. Which two?
- a) The cluster assignment step, where the parameters $C(i)$ are updated.
 - b) Move the cluster centroids, where the centroids μ_k are updated.
 - c) Feature scaling, to ensure each feature is on a comparable scale to the others.
 - d) Using the elbow method to choose K .

The correct choice a and b

a.....The assignment step is the first step of the K-means loop.

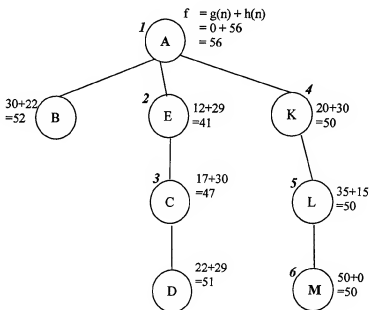
b.....The cluster update is the second step of the K-means loop.

Question No. 2

Question No. 3

Question 1 – Model Answer

a) The search tree is produced, showing the cost function at each node
(The numbers in *italic* show the order in which the nodes were expanded)



I will award marks on the following basis

- 6 marks : for getting the search tree correct
- 3 marks : for showing separate (correct) $f(n)$ and $g(n)$ values (i.e. separating the two and showing them correctly)
- 3 marks : for showing the correct total values for each node
- 2 marks : for showing the order in which the nodes were expanded (these are shown as the numbers in *italic*)

The marks will be awarded pro-rata, if applicable.

State the order in which the nodes were expanded

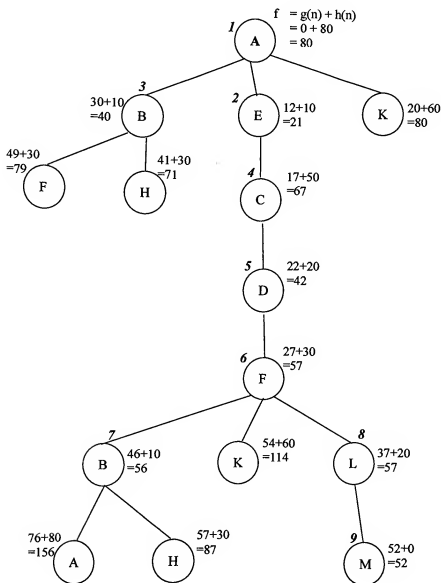
The route is A, E, C, K, L, M (1 mark)
The cost is 50 (1 mark)

Note that E and C are expanded before K. If this is not done (by providing the answer of A, K, L, M) then the student has not applied the algorithm correctly.

- b) What route would now be returned by the A* algorithm and what would the cost of that route be?**
The answer I am looking for is
The route is A, E, B, C, D, F, B, L, M (1 mark)
The cost is 52 (1 mark)

G5AIAI – Introduction to AI Exam
2004/2005

This part of the question does not ask for the search tree but the students will have to draw it in order to answer the question (the search tree is shown below). Therefore, 4 marks will be awarded for drawing the search tree (the marks will be awarded pro-rata, depending on how much of the tree they get right).



c) How do you account for the different routes returned by the two A* algorithms?

I am looking for the following statements from the student.

- The first table was an **admissible heuristic**, that is one that **never over-estimates** the cost to the goal (1 mark)
- The second table was not an admissible heuristic in that some of the values did over-estimate the cost to the goal (1 mark).
- The A* algorithm is only guaranteed to find the optimal solution **if the heuristic used is admissible** (1 mark).

G5AIAI – Introduction to AI Exam
2004/2005

It is *not* permissible for the student to simply state “*because the straight line heuristic cost table has changed, the A* algorithm will obviously return a different solution.*” I am looking for an understanding of the concept of admissibility.

Note : Although the students have no proof that one heuristic is admissible and the other is not, this is obvious as one heuristic returned the optimal solution and the other did not.